

TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results

Group: (SetuServ) SetuServ

Run ID: SSEC3inclML

Run type: Primary

Description of run:

Post Error Correction (likely errors are identified using M/L), counting “not sures” with ML score >50% as relevant docs & remaining “not sures” as irrelevant

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	21	1983	46	6	0.768	0.977	0.023	0.232	0.078	0.968
416	1235	45	40	1184	6	5	0.880	0.995	0.005	0.120	0.027	0.965
417	2992	75	59	2835	82	16	0.783	0.972	0.028	0.217	0.082	0.890
420	1136	37	33	1056	43	4	0.882	0.960	0.040	0.118	0.069	0.960
427	1528	37	31	1480	11	6	0.829	0.992	0.008	0.171	0.038	0.956
432	2503	22	9	2417	64	13	0.413	0.974	0.026	0.587	0.163	0.712
438	1798	162	152	1514	122	10	0.936	0.925	0.075	0.064	0.069	0.951
445	1404	60	49	1327	17	11	0.811	0.987	0.013	0.189	0.052	0.935
446	2020	156	96	1815	49	60	0.615	0.973	0.027	0.385	0.116	0.834
447	1588	16	14	1572	0	2	0.853	1.000	0.000	0.147	0.007	0.968
Average	1826.000	63.700	50.400	1718.300	44.000	13.300	0.777	0.976	0.024	0.223	0.070	0.914

Table 1: This table shows per-topic statistics and overall averages for the run SSEC3inclML. The topics are 10 randomly selected topics from the TREC 8 ad-hoc task. A relevant document is positive and a non-relevant document is negative. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts are based on an adjudicated set of relevance judgments that differs from the original TREC-8 ad-hoc qrels. The true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and the false negative rate (FNR) are all smoothed values. Details of the computation of the logistic average misclassification (LAM) rate and the area under the curve (AUC) are given in the track overview paper. Some runs did not report a probability of relevance and thus will have NA for their AUC score.

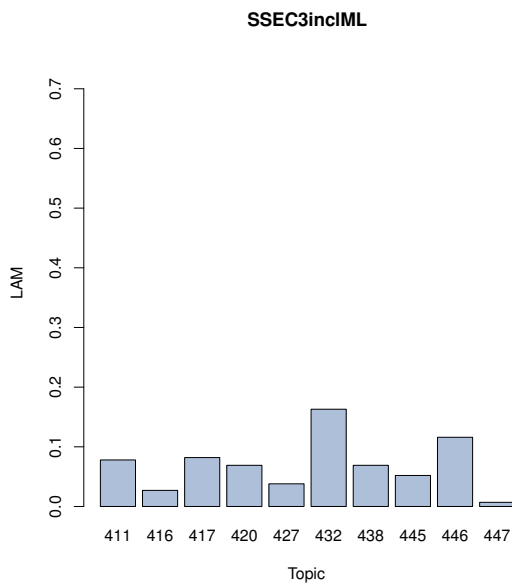


Figure 1: SSEC3inclML LAM

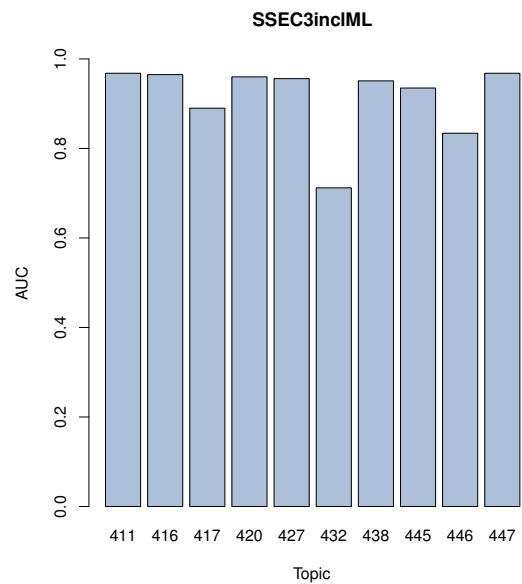


Figure 2: SSEC3inclML AUC